

アイヌ語のテキストマイニング

市立札幌旭丘高校数理データサイエンス科 ヒンナコタン 2年 木嶋もも菜 佐藤優空 藤村結愛 松本結愛

【要約】北海道の先住民として知られるアイヌ民族は、文字を持たない民族である。また、日本とは異なった自然観を持ち、独自の文化を形成している。そこで、私たちは口承されてきたアイヌ民族の民譚をテキストマイニングすることで傾向を読み取ろうとした。その結果、アイヌ文献は日本昔話と比べて人称名詞が少ないこと、ウエペケレはカムイユカラと比べて動詞の種類が少なく名詞が多いことなどが分かった。

1. 研究内容

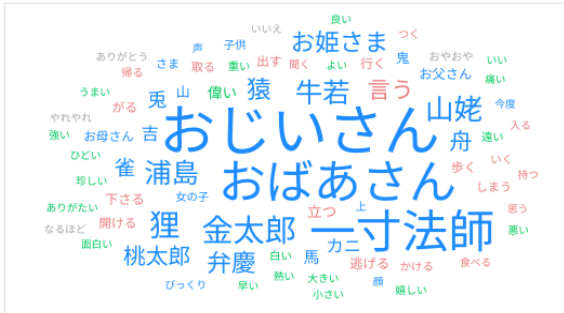
アイヌ語の文献をテキストマイニングし、特徴や傾向を調べた。日本昔話とアイヌ文献の比較、ウエペケレ(散文)とカムイユカラ(韻文)の比較から、それぞれの特徴を分析した。

2. 研究手法

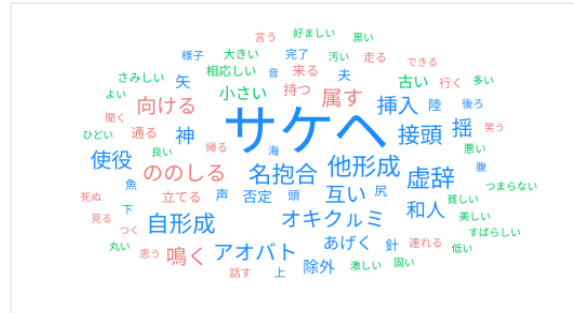
国立国語研究所が運営しているアイヌ語口承文芸コーパスに、38作のアイヌ文献が掲載されている。そのうちの計29作の逐語訳データを収集した。また、青空文庫より、日本昔話を12作収集した。それらを、ユーザーローカルが提供しているAIテキストマイニングツールを利用し、異なる2種類の文書を解析・比較することで、傾向を調べた。

3. 結果と考察

日本語



アイヌ



日本語	アイヌ文献
人称の名詞が多い	人称の名詞が少ない
ウエペケレ	カムイユカラ
動詞の種類・頻度が少ない 名詞が多い	動詞の種類・頻度が多い 名詞が少ない

名詞			動詞		
ウエペケレ	単語	カムイユカラ	ウエペケレ	単語	カムイユカラ
0	サケへ	100	53	行く	47
89	他形成	11	37	持つ	63
97	挿入	3	35	来る	65
90	名抱合	10	26	言う	74
91	声	9	55	帰る	45
100	母	0	39	向ける	61
100	小屋	0	39	見る	61
100	父	0	45	思う	55
100	狩	0	0	属す	100
86	神	14	0	住む	100
88	自形成	12	24	話す	76
99	山	1	47	遅れる	53
91	否定	9	7	聞く	93
99	子音	1	0	鳴く	100
100	塊	0	9	作る	91
100	少年	0	29	死ぬ	71
100	糸	0	77	食べる	23
0	虚辞	100	30	走る	70

日本語とアイヌ文献の比較より、日本語は三人称のものが多く、アイヌの物語は一人称のものが多くと推測した。実際にアイヌ文献には、伝聞の形が取られているものが多くあった。アイヌ文献が1人称で多く語られるのは、3人称視点であると複雑になってしまう点、また、口承が前提である点にあるのではないかと考察した。

また、ウエペケレとカムイユカラの比較より、韻文であるカムイユカラは音節が決まっていることから、文の構造が散文のウエペケレと比べて単純になっていると推測した。カムイユカラは主語動詞の優先順位が高くなり、動詞の種類や頻度が増加する。また、ウエペケレは自由な文章を作成できるため、修飾などが増え、それに伴い名詞も増加する。このような関係にあると考察した。